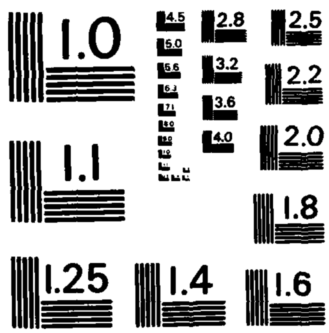
AD-A121 143   SELECTION OF VARIABLES IN DISCRIMINANT ANALYSIS(U)   1/1
PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS
P R KRISHNAIAH JUN 82 TR-82-15 AFOSR-TR-82-0944

UNCLASSIFIED   F49629-82-K-0001   F/G 12/1   NL

END
FILMED
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5

Center for Multivariate Analysis

University of Pittsburgh

SELECTION OF VARIABLES IN
DISCRIMINANT ANALYSIS**

P. R. Krishnaiah*

University of Pittsburgh

June 1982

Technical Report No. 82-15

Center for Multivariate Analysis
University of Pittsburgh
Ninth Floor, Schenley Hall
Pittsburgh, PA 15260

A

*The work of the author is sponsored by the Air Force Office of Scientific Research under Contract No. F49629-82-K-0001. Reproduction in whole or in part is permitted for any purpose of the United States Government.

**This paper will appear in Volume 2 of Handbook of Statistics.

1. INTRODUCTION

In a number of disciplines, data analysts are confronted with the problem of classifying an observation into one of the distinct groups when the number of variables is very large. So, it is of interest to find out a smaller number of important variables which are adequate for discrimination. These variables maybe a subset of the original variables or certain linear combinations of the original variables. The selection of variables is important since there are situations where inclusion of unimportant variables may actually decrease the ability for discrimination. Apart from it, it is more feasible to analyze the data from cost and computational considerations if the number of variables is small. In this chapter, we discuss various procedures for the selection of variables in discriminant analysis.

In Section 2, we discuss procedures to find out whether certain discriminant coefficients associated with variables are important for discrimination between two populations. In Section 3, generalizations of the above procedures for several populations are discussed. The problems of testing the hypotheses on discriminant coefficients using simultaneous test procedures will be discussed in another paper. The procedures in Sections 2 and 3 are based upon using conditional distributions. In Section 4 we discuss various procedures to determine the number of important discriminant functions.

2. TESTS ON DISCRIMINANT FUNCTIONS USING CONDITIONAL DISTRIBUTIONS FOR TWO POPULATIONS

In this section, we discuss procedures for testing the hypotheses on the coefficients of the discriminant functions associated with the discrimination between two multivariate normal populations. Let the mean vector and covariance matrix of i th multivariate normal populations be given by $\underset{\sim}{\mu}_i$ and Σ. Now, consider the discriminant function $\underset{\sim}{a}'\underset{\sim}{x}$ for the two population case where $\underset{\sim}{a}' = (a_1,\ldots,a_p) = (\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)'\Sigma^{-1}$ and $\underset{\sim}{x}' = (x_1,\ldots,x_p)$. If any of the coefficients a_i are zero, then the corresponding variables x_i do not make any contribution for the discrimination between the two populations. So, it is of interest to find out as to which of the coefficients are zero. Suppose, we know <u>a priori</u> that $x_1,\ldots,x_q$ are important and we are not sure of $x_{q+1},\ldots,x_p$. Then, we are interested in testing the hypothesis that $a_{q+1} = \ldots = a_p = 0$.

Let $\underset{\sim}{x}_1' = (x_1,\ldots,x_q)$, $\underset{\sim}{x}_2' = (x_{q+1},\ldots,x_p)$, $\underset{\sim}{\delta}' = (\delta_1,\ldots,\delta_p)$ and let $\underset{\sim}{\mu}_1,\underset{\sim}{\mu}_2,\Sigma$ be partitioned as

$$\underset{\sim}{\mu}_i = \begin{pmatrix} \underset{\sim}{\mu}_{i1} \\ \underset{\sim}{\mu}_{i2} \end{pmatrix}, \quad \underset{\sim}{\delta} = \begin{pmatrix} \underset{\sim}{\delta}_1 \\ \underset{\sim}{\delta}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

and $\underset{\sim}{a}' = (\underset{\sim}{a}_1',\underset{\sim}{a}_2')$. Also, $\underset{\sim}{\mu}_{i1},\underset{\sim}{a}_1,\underset{\sim}{\delta}_1$ are of order $q\times 1$ and Σ_{11} is of order $q\times q$. Let H denote the hypothesis that the conditional mean vector of $\underset{\sim}{x}_2$ given $\underset{\sim}{x}_1$ is the same for both populations. Also, let $\beta = \Sigma_{21}\Sigma_{11}^{-1}$. Then

$$H: \underset{\sim}{\mu}_{12}-\beta\underset{\sim}{\mu}_{11} = \underset{\sim}{\mu}_{22}-\beta\mu_{21}. \qquad (2.1)$$

Now, let $\Delta = (\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)' \Sigma^{-1}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)$, $\Delta_1 = (\underset{\sim}{\mu}_{11}-\underset{\sim}{\mu}_{21})' \Sigma_{11}^{-1}(\underset{\sim}{\mu}_{11}-\underset{\sim}{\mu}_{21})$ and $\Delta_{2\cdot 1} = (\underset{\sim}{\delta}_2-\beta\underset{\sim}{\delta}_1)' \Sigma_{2\cdot 1}^{-1}(\underset{\sim}{\delta}_2-\beta\underset{\sim}{\delta}_1)$. The hypothesis H is equivalent to $\underset{\sim}{\delta}_2-\beta\underset{\sim}{\delta}_1 = \underset{\sim}{0}$. It is known (see Rao (1946)) that H is equivalent to the hypothesis that $a_{q+1} =\ldots= a_p = 0$ and it is also equivalent to the hypothesis that $\Delta = \Delta_1$. This can be interpreted as H is equivalent to the hypothesis that the distance between the two populations based on the p variables is equal to the distance between the populations on the basis of the first q variables.

Let $\underset{\sim}{x}_i' = (x_{i1},\ldots,x_{iq},x_{iq+1},\ldots,x_{ip}) = (\underset{\sim}{z}_{i1}',\underset{\sim}{z}_{i2}')$, $(i = 1,2)$, be distributed as multivariate normal with mean vector $\underset{\sim}{\mu}_i (i = 1,2)$ and covariance matrix Σ where $\underset{\sim}{z}_{i1}$ is of order $q\times 1$. Then, the conditional distribution of $\underset{\sim}{z}_{i2}$ given $\underset{\sim}{z}_{i1}$ is multivariate normal with covariance matrix $\Sigma_{2\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ and mean vector $\underset{\sim}{\eta}_i + \beta\underset{\sim}{z}_{i1}$ where $\underset{\sim}{\eta}_i = \underset{\sim}{\mu}_{i2} - \beta\underset{\sim}{\mu}_{i1}$. We wish to test $\underset{\sim}{\eta}_1 = \underset{\sim}{\eta}_2$, that is, $\underset{\sim}{\delta}_2 - \beta\underset{\sim}{\delta}_1 = 0$. This can be done by using any of the known procedures. Let $(\underset{\sim}{z}_{i1t}',\underset{\sim}{z}_{i2t}')$, $(t = 1,2,\ldots,n_i)$, be t-th observation on $(\underset{\sim}{z}_{i1}',\underset{\sim}{z}_{i2}')$. Then, the conditional model is given by

$$E_c(\underset{\sim}{z}_{i2t}/\underset{\sim}{z}_{i1t}) = \underset{\sim}{\eta}_i + \beta\underset{\sim}{z}_{i1t} \tag{2.2}$$

for $i = 1,2$ and $t = 1,2,\ldots,n_i$.

We can test the hypothesid H by using the following statistic:

$$F = \frac{c(\hat{\underset{\sim}{\delta}}_2' - \hat{\beta}\hat{\underset{\sim}{\delta}}_1)' S_{e2\cdot1}^{-1}(\hat{\underset{\sim}{\delta}}_2 - \hat{\beta}\hat{\underset{\sim}{\delta}}_1)(n-p-1)}{(1+c\hat{\underset{\sim}{\delta}}_1' S_{e11}^{-1}\hat{\underset{\sim}{\delta}}_1)(p-q)} \tag{2.3}$$

where $S_{e2\cdot1} = S_{e22} - S_{e21}S_{e11}^{-1}S_{e12}$, $\hat{\beta} = S_{e21}S_{e11}^{-1}$, $\hat{\delta}_j = \bar{\underset{\sim}{z}}_{j1\cdot} - \bar{z}_{j2\cdot}$, $n_j\bar{\underset{\sim}{z}}_{jk\cdot} = \sum_{t=1}^{n_j} \underset{\sim}{z}_{jkt}$, $n = n_1 + n_2$, $c = n_1n_2/n$ and

$$S = \begin{pmatrix} S_{e11} & S_{e12} \\ S_{e21} & S_{e22} \end{pmatrix} = \sum_{i=1}^{2}\sum_{t=1}^{n_i} \begin{pmatrix} \underset{\sim}{z}_{i1t} - \bar{\underset{\sim}{z}}_{i1\cdot} \\ \underset{\sim}{z}_{i2t} - \bar{\underset{\sim}{z}}_{i2\cdot} \end{pmatrix} (\underset{\sim}{z}_{i1t} - \bar{\underset{\sim}{z}}_{i1\cdot}, \underset{\sim}{z}_{i2t} - \bar{\underset{\sim}{z}}_{i2})'.$$

The statistic F is distributed as the central F distribution with (p-q,n-p-1) degrees of freedom when H is true. The hypothesis H is accepted or rejected according as

$$F \lessgtr F_\alpha \tag{2.5}$$

where

$$P[F \le F_\alpha | H] = (1-\alpha). \tag{2.6}$$

The above procedure for testing the hypothesis $a_{q+1} = \ldots = a_p = 0$ was proposed by Rao (1946, 1966). It is known (e.g., see Kshirsagar (1972)) that Rao's U statistic is related to the F statistic given in (2.3).

The simultaneous confidence intervals associated with the above procedure are known to be

$$|b'(\hat{\underset{\sim}{\delta}}_2 - \hat{\beta}\hat{\underset{\sim}{\delta}}_1 - \underset{\sim}{\delta}_2 + \beta\underset{\sim}{\delta}_1)| \le \sqrt{F_\alpha \underset{\sim}{b}' S_{e2\cdot1}\underset{\sim}{b}\ (p-q)(1+c\hat{\underset{\sim}{\delta}}_1' S_{e11}^{-1}\hat{\underset{\sim}{\delta}}_1)/c(n-p-1)} \tag{2.7}$$

for all nonnull $\underset{\sim}{b}$.

3. TESTS ON DISCRIMINANT FUNCTIONS FOR SEVERAL POPULATIONS USING CONDITIONAL DISTRIBUTIONS

In this section, we consider the problem of testing the hypotheses that the discriminant coefficients associated with certain variables in the discriminant functions are zero. Let $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_k$ be distributed independently as multivariate normal with mean vectors $\underset{\sim}{\mu}_1,\ldots,\underset{\sim}{\mu}_k$ and covariance matrix Σ. Also, let $\underset{\sim}{x}_{ij}$, $(j=1,2,\ldots,n_i)$, denote j-th independent observation on $\underset{\sim}{x}_i$. Then, the between group sums of squares and cross products (SP) matrix is given by

$$S = \sum_{i=1}^{k} n_i(\bar{\underset{\sim}{x}}_{i.}-\bar{\underset{\sim}{x}}_{..})(\bar{\underset{\sim}{x}}_{i.}-\bar{\underset{\sim}{x}}_{..})'$$

$$= \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

where

$$n_i\bar{\underset{\sim}{x}}_{i.} = \sum_j x_{ij},\ n = n_1+\ldots+n_k,\ n\bar{\underset{\sim}{x}}_{..} = \sum_i \sum_j \underset{\sim}{x}_{ij}.$$

Now, let $\theta_1 \geq \ldots \geq \theta_p$ denote the eigenvalues of the noncentrality matrix $\Omega = \Delta\Sigma^{-1}$ and let $\underset{\sim}{\nu}_i' = (\nu_{i1},\ldots,\nu_{ip})$, $(i=1,\ldots,p)$, denote the eigenvector corresponding to θ_i where

$$\Delta = \sum_{i=1}^{k} n_i(\underset{\sim}{\mu}_i-\bar{\underset{\sim}{\mu}})(\underset{\sim}{\mu}_i-\bar{\underset{\sim}{\mu}})' \tag{3.1}$$

and $\bar{\underset{\sim}{\mu}} = (n_1\underset{\sim}{\mu}_1+\ldots+n_k\underset{\sim}{\mu}_k)/n$. Suppose the rank of Ω is r. Then, $\theta_{r+1}=\ldots=\theta_p = 0$ and we have r meaningful discriminant functions. The within group SP matrix is given by

$$S_e = \sum_{i}\sum_{j}(\underset{\sim}{x}_{ij}-\bar{\underset{\sim}{x}}_{i.})(\underset{\sim}{x}_{ij}-\bar{\underset{\sim}{x}}_{i.})' = \begin{pmatrix} S_{e11} & S_{e12} \\ S_{e21} & S_{e22} \end{pmatrix}$$

where S_{e11} is of order $q \times q$. Now, let us partition $\underset{\sim}{\mu}_i$, $\underset{\sim}{\nu}_j$ and Σ as follows:

$$\underset{\sim}{\mu}_i = \begin{pmatrix} \underset{\sim}{\mu}_{i1} \\ \underset{\sim}{\mu}_{i2} \end{pmatrix}, \qquad \underset{\sim}{\nu}_j = \begin{pmatrix} \underset{\sim}{\nu}_{j1} \\ \underset{\sim}{\nu}_{j2} \end{pmatrix},$$

$$\Delta = \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\underset{\sim}{\mu}_{i1}$ and $\underset{\sim}{\nu}_{j1}$ are of order $q \times 1$, and Δ_{11} and Σ_{11} are of order $q \times q$. Let H_j: $\underset{\sim}{\nu}_{j2} = 0$ for $j = 1,2,\ldots,r$ and $H = \bigcap_{j=1}^{r} H_j$. It is known (e.g. see McKay (1977) and Fujikoshi (1980)) that the hypothesis H and the following statements are equivalent:

$$\underset{\sim}{\mu}_{12} - \Sigma_{21}\Sigma_{11}^{-1}\underset{\sim}{\mu}_{11} = \ldots = \underset{\sim}{\mu}_{k2} - \Sigma_{21}\Sigma_{11}^{-1}\underset{\sim}{\mu}_{k1} \tag{3.2}$$

$$\mathrm{tr}\Delta\Sigma^{-1} = \mathrm{tr}\Delta_{11}\Sigma_{11}^{-1}. \tag{3.3}$$

The hypothesis given by (3.2) can be tested by using various known methods (e.g., see Kshirsagar (1972) and Rao (1973)) like Roy's largest root test, likelihood ratio test, etc.

4. TESTS FOR THE NUMBER OF IMPORTANT DISCRIMINANT FUNCTIONS

It is well known that Fisher's linear discriminant functions are the best for discrimination among all linear functions of the original variables. In this section, we review some procedures for the selection of important discriminant functions.

We know that

$$\frac{(n-k-p-1)}{(k-1)} E(S\, S_e^{-1}) = I + \frac{\Omega}{(k-1)} = \Sigma^* \qquad (4.1)$$

Let $\lambda_1 \geq \ldots \geq \lambda_p$ be the eigenvalues of Σ^*. Then $\lambda_i = 1 + (\theta_i/(k-1))$ where θ_i was defined in Section 3. The p discriminant functions are $\underset{\sim}{\nu}_1'x, \ldots, \underset{\sim}{\nu}_p'\underset{\sim}{x}$ where $\underset{\sim}{x}$ is a vector of observations on the p variables. $\underset{\sim}{\nu}_i'x$ is i-th most important discriminant function and $\underset{\sim}{\nu}_i$ was defined in Section 3. The problem of testing for the rank of the noncentrality matrix Ω is equivalent to the problem of testing for the number of important discriminant functions. It is also equivalent to testing the following hypothesis on certain structural relations among the components of the mean vectors:

$$A\underset{\sim}{\mu}_i = \xi, \; (i = 1,2,\ldots,k) \qquad (4.2)$$

where A: s×p and $\underset{\sim}{\xi}$: p×1 are unknown and the rank of A is s. The above hypothesis implies that the points $\underset{\sim}{\mu}_1, \ldots, \underset{\sim}{\mu}_k$ lie in a r-dimensional space where r = p − s.

Now, let H_i denote the hypothesis that $\lambda_i = 1$. Then the hypothesis (4.2) is equivalent to the hypothesis that $\bigcap_{i=r+1}^{p} H_i$ is true. The problem of testing for the rank of Ω was originally considered by Fisher (1938). Fisher's test to test the hypothesis that the rank of Ω is r is based upon the statistic

$$T_1 = (\ell_{r+1}+\ldots+\ell_p) \tag{4.3}$$

where $\ell_{r+1},\ldots,\ell_p$ are the eigenvalues of SS_e^{-1}.

We can test for the rank of Ω within the framework of simultaneous test procedures as follows. Accept or reject $H_i\,(i=r+1,\ldots,p)$ according as

$$\ell_i \lesseqgtr c_\alpha \tag{4.4}$$

where

$$P[\ell_{r+1} \leq c_\alpha | H_{r+1}] = (1-\alpha). \tag{4.5}$$

If H_t is accepted but H_{t-1} is rejected, then the rank of Ω is $(t-1)$. But, the distribution of ℓ_{r+1} involves $\lambda_1,\ldots,\lambda_r$ as nuisance parameters even when H_{r+1} is true and so the exact values of c_α cannot be computed. If we know in advance that $H_t,\ldots,H_p$, $(t > r+1)$, are true then we test $H_{r+1},\ldots,H_{t-1}$ only. In many situations, it is of interest to test $H_1,\ldots,H_p$ simultaneously since we don't know in advance that H_r is not true. In these situations, we accept or reject H_i according as

$$\ell_i \lesseqgtr c_{\alpha 1} \tag{4.6}$$

where

$$P[\ell_1 \leq c_{\alpha 1} | H_1] = (1-\alpha). \tag{4.7}$$

Here, we note that the hypotheses $H_1, \ldots, H_p$ are nested. For example, H_i implies $H_{i+1}, \ldots, H_p$. When H_1 is true, the exact distribution of ℓ_1 was given in Krishnaiah and Chang (1971). For percentage points of the distribution of ℓ_1 in the null case, the reader is referred to Krishnaiah (1980) and Pillai (1960). A review of the literature on the distributions of individual roots and certain functions of the eigenvalues is given in Krishnaiah (1978) and Muirhead (1978).

Fang and Krishnaiah (1982) derived asymptotic nonnull distributions of certain functions of the eigenvalues of some random matrices when the underlying distribution is not normal.

The likelihood ratio statistic for testing the hypothesis H_{r+1} is known to be

$$L_1 = \prod_{i=r+1}^{p} (1+\ell_i)^{-n/2}. \tag{4.8}$$

For large samples, $-2\log L_1$ is distributed approximately as $n(\ell_{r+1}+\ldots+\ell_p) = nT_1$. It is known (see Hsu (1941a)) that nT_1 is distributed asymptotically as chi-square with $(p-r)(k-1-r)$ degrees of freedom when n_i's tend to infinity. Bartlett (1947) showed that $T_2 = C_1 \sum_{i=r+1}^{p} \log(1+\ell_i)$ is distributed as chi-square with $(p-r)(k-1-r)$ degrees of freedom where the correction factor C_1 is given by $C_1 = (n-1-(p+k)/2)$. The chi-square approximation to the distribution of the statistic T_2 is better that the corresponding approximation to nT_1. Lawley (1959) suggested a modified correction factor but it involves nuisance parameters. For further details on the likelihood ratio test for the rank of Ω, the reader is referred to Kshirsagar (1972) and Rao (1973). Anderson (1951b) derived the likelihood ratio test statistic for testing the rank of regression matrix under the multivariate regression model.

In general, we test the hypothesis H_{r+1} by using $\psi(\ell_{r+1},\ldots,\ell_p)$, a suitable function of $\ell_{r+1},\ldots,\ell_p$, as follows. Accept or reject H_{r+1} according as

$$\psi(\ell_{r+1},\ldots,\ell_p) \lesseqgtr c_{\alpha 2} \tag{4.9}$$

where

$$P[\psi(\ell_{r+1},\ldots,\ell_p) \leq c_{\alpha 2} | H_{r+1}] = (1-\alpha). \tag{4.10}$$

Some special cases of $\psi(\ell_{r+1},\ldots,\ell_p)$ are $\ell_{r+1}, (\ell_{r+1}+\ldots+\ell_p)$, etc.

We now review some known results on asymptotic distributions of ℓ_i's and certain functions of these eigenvalues.

Let $\ell_1 \geq \ldots \geq \ell_v$, $(r \leq v \leq p)$, be the nonzero eigenvalues of SS_e^{-1}. Also, let the eigenvalues of Ω have multiplicities as below.

$$\begin{array}{lcll}
\theta_1 & =\ldots= & \theta_{p_1^*} & = n\delta_1 \\
\theta_{p_1^*+1} & =\ldots= & \theta_{p_2^*} & = n\delta_2 \\
\vdots & \vdots & \vdots & \\
\theta_{p_{t-1}^*+1} & =\ldots= & \theta_{p_t^*} & = n\delta_t \\
\theta_{p_t^*+1} & =\ldots= & \theta_p & = 0 \quad ,
\end{array} \tag{4.11}$$

where $p_j^* = p_1+\ldots+p_j$, $(j=1,2,\ldots,t+1)$, $r=p_t^*$, $v = p_{t+1}^*$, and $p_o^* = 0$. In addition, let

$$u_{i_h} = \sqrt{n}\ (2\delta_h^2 + 4\delta_h)^{-\frac{1}{2}}(\ell_{i_h} - \delta_h)$$
$$u_{r+j} = n\ell_{r+j} \tag{4.12}$$

where $h = 1,2,\ldots,t$, $i_h = p_{h-1}^*+1,\ldots,p_h^*$ and $j = 1,\ldots,v-r$. Also, let $n_i = n_o q_i$ for $i = 1,2,\ldots,k$. Then, the limiting distribution of $u_1,\ldots,u_v$, as $n_o\to\infty$, derived by Hsu (1941b) is given by

$$f(u_1,\ldots,u_v) = \prod_{j=1}^{t+1} \eta_j(u_{p_{j-1}^*+1},\ldots,u_{p_j^*}) \tag{4.13}$$

where $\eta_j(\cdot)(j = 1,2,\ldots,t)$, denotes the joint density of the eigenvalues of A_j and the elements of A_j: $p_j^* \times p_j^*$ are distributed independently as normal with mean zero. The variances of the diagonal elements of A_j are equal to one whereas the variances of the off-diagonal elements are equal to $\frac{1}{2}$. Also, $\eta_{t+1}(\cdot)$ is the joint density of the eigenvalues of A_{t+1}: $(v-r)\times(v-r)$ where A_{t+1} is distributed as the central Wishart matrix with $(k-1-r)$ degrees of freedom and $E(A_{t+1}) = (k-r-1)I_{v-r}$. Here $A_1,\ldots,A_{t+1}$ are distributed independent of each other. Expressions for the densities of the eigenvalues of A_j $(j = 1,2,\ldots,t)$, and A_{t+1} were given in Hsu (1941b). The asymptotic joint density of $u_1,\ldots,u_v$ given by (4.13) was derived by Anderson (1951a) by a different method.

Now let $\theta_i = m\beta_i$, $(i=1,2,\ldots,r)$, $\theta_{r+1}=\ldots=\theta_p=0$ where m is a suitable correction factor and β_i's are constants. Then, Fujikoshi (1976) derived approximations to the distributions of $m_i T_i$, $(1=1,2,3)$ up to terms of order m_i^{-2} where

$$T_1 = \sum_{j=r+1}^{p} \log(1+\ell_j) \tag{4.14}$$

$$T_2 = \sum_{j=r+1}^{p} \ell_j \tag{4.15}$$

$$T_3 \quad \sum_{j=r+1}^{p} \{\ell_j/(1+\ell_j)\} \tag{4.16}$$

and m_1, m_2 and m_3 are suitable correction factors. The first terms in these approximations involve chi-square distribution. Similar approximations can be derived for various other functions of $\ell_{r+1},\ldots,\ell_p$. Asymptotic distributions of a wide-class of functions of $\ell_1,\ldots,\ell_p$ in the nonnull cases were given in Fujikoshi (1978), Krishnaiah and Lee (1979) and Fang and Krishnaiah (19

In some situations, we know in advance that the last few eigenvalues of Ω are equal to zero. For example, when $p > k-1$, $\theta_k = \theta_{k+1}=\ldots=\theta_p = 0$. In these situations, it is of interest to test whether some of the θ_i's$(i=1,2,\ldots,k-1)$ are zero. We can test the hypotheses $H_j(j=t,t+1,\ldots,k-1)$ as follows also. We accept or reject $H_j(j=t,t+1,\ldots,k-1)$ according as

$$T_{jk} \lesseqgtr c_\alpha \tag{4.17}$$

where

$$P[T_{jk} \leq c_\alpha;\ j=t,\ldots,k-1 \mid \bigcap_{j=t}^{k-1} H_j] = (1-\alpha), \tag{4.18}$$

$T_{jk} = \ell_j/(\ell_k+\ldots+\ell_p)$. As pointed out earlier, the joint distribution of $\ell_t, \ell_{t+1}, \ldots, \ell_p$, when H_t is true and the sample sizes tend to infinity, is the same as the joint distribution of the eigenvalues of the central Wishart matrix. Exact distribution of the ratio of the largest root to the sum of the roots of the central Wishart matrix was considered in Schuurmann, Krishnaiah and Chattopadhyay (1973) and Krishnaiah and Schuurmann (1974).

We now discuss the problems of testing the hypotheses $H_1, \ldots, H_p$ in an ad hoc sequential way using conditional distributions. The hypothesis H_1 is accepted or rejected according as

$$\ell_1 \lesseqgtr c_{\alpha 1} \tag{4.19}$$

where

$$P[\ell_1 \leq c_{\alpha 1} \mid H_1] = (1-\alpha_1). \tag{4.20}$$

If H_1 is accepted, we conclude that $\Omega = 0$ and don't proceed further. If H_1 is rejected, we accept or reject H_2 according as

$$\ell_2 \lesseqgtr c_{\alpha 2} \tag{4.21}$$

$$P[\ell_2 \leq c_{\alpha 2} \mid \ell_1 \geq c_{\alpha 1}; H_2] = (1-\alpha_2). \tag{4.22}$$

If H_2 is accepted, we don't proceed further. Otherwise we accept or reject H_3 according as

$$\ell_3 \lessgtr c_{\alpha 3}$$

where

$$P[\ell_3 \leq c_{\alpha 3} | \ell_1 \geq c_{\alpha 1},\ \ell_2 \geq c_{\alpha 2}; H_3] = (1-\alpha_3). \tag{4.23}$$

In general, if we accept H_i, we don't proceed further. Otherwise, we accept or reject H_{i+1} according as

$$\ell_{i+1} \lessgtr c_{\alpha,i+1}$$

where

$$P[\ell_{i+1} \leq c_{\alpha,i+1} | \ell_j \geq c_{\alpha j},\ j=1,2,\ldots,i; H_{i+1}] = (1-\alpha_{i+1}). \tag{4.24}$$

Then, the overall type I error to test $H_1, H_2, \ldots, H_{i+1}$ sequentially is given by α^*_{i+1} where

$$\prod_{t=0}^{i} P[\ell_{t+1} \leq c_{\alpha,t+1} | \ell_j \geq c_{\alpha j};\ j=1,2,\ldots,t; H_{t+1}] = (1-\alpha^*_{i+1}) \tag{4.25}$$

Chou and Muirhead (1979) derived asymptotic conditional distribution of ℓ_{t+1} given $\ell_1,\ldots,\ell_t$. This distribution involves $\theta_1,\ldots,\theta_t$ as nuisance parameters even when H_{t+1} is true. If we ignore the terms of order $\theta_i^{-1}, (i=1,2,\ldots,t)$, then the conditional distribution of ℓ_{t+1} given $\ell_1,\ldots,\ell_t$ does not involve nuisance parameters when H_{t+1} is true. If we further ignore the linkage factors

$$\prod_{i=1}^{t} \prod_{j=t+1}^{p} (\ell_i - \ell_j)^{\frac{1}{2}}$$

the joint distribution of the latent roots $\ell_{t+1},\ldots,\ell_p$ is the same as the joint distribution of $S_1(S_1+S_2)^{-1}$ where S_1 and S_2 are distributed independently as central Wishart matrices with (k-1-t) and (n-k-t) degrees of freedom respectively and $E\{S_1/(k-1-t) = E(S_2/(n-k-t))\} = I_{p-t}$. In deriving the above result, Chou and Muirhead assumed that the error degrees of freedom is very large but the noncentrality matrix remain fixed. For a discussion of the asymptotic conditional distribution of $\ell_{t+1},\ldots,\ell_p$ given $\ell_1,\ldots,\ell_t$ under some other conditions, the reader is referred to Muirhead (1978).

REFERENCES

Anderson, T. W. (1951a). The asymptotic distribution of certain characteristic roots and vectors. In J. Neyman, ed., Proceedings of the Second Berkeley Symposium in Mathematical Statistics and Probability, pp. 103-130. University of California Press, Berkeley.

Anderson, T. W. (1951b). Estimating linear restrictions on regression coefficients for multivariate normal distributions. Ann. Math. Statist., **22**, 327-351.

Bartlett, M. S. (1947). Multivariate analysis. J. Roy. Statist. Soc. Suppl., **9**, 176-190.

Chou, R. J. and Muirhead, R. J. (1979). On some distribution problems in MANOVA and discriminant analysis. J. Multivariate Anal., **9**, 410-419.

Fang, C. and Krishnaiah, P. R. (1981). Asymptotic distributions of functions of the eigenvalues of the real and complex noncentral Wishart matrices. In Csorgo, M., Dawson, D. A., Rao, J. N. K. and Saleh, A.K. Md. E., eds., Statistics and Related Topics. North-Holland Publishing Company.

Fang, C. and Krishnaiah, P. R. (1982). Asymptotic distributions of functions of the eigenvalues of some random matrices for nonnormal populations. J. Multivariate Anal. **12**, 39-63.

Fisher, R. A. (1938). The statistical utilization of multiple measurements. Ann. Eugen., **8**, 376-386.

Fujikoshi, Y. (1976). Asymptotic expressions for the distributions of some multivariate tests. In P. R. Krishnaiah, ed., Multivariate Analysis - IV, pp. 55-71. North-Holland Publishing Company.

Fujikoshi, Y. (1978). Asymptotic expansions for the distributions of some functions of the latent roots of matrices in three situations. J. Multivariate Anal., **8**, 63-72.

Fujikoshi, Y. (1980). Tests for additional information in canonical discrimination analysis and canonical correlation analysis. Technical Rept. No. 12, Statistical Research Group, Hiroshima University.

Hsu, P. L. (1941a). On the problem of rank and the limiting distribution of Fisher's test functions. Ann. Eugenics, **11**, 39-41.

Hsu, P. L. (1941b). On the limiting distribution of roots of a determinantal equation. J. London Math. Soc., **16**, 183-194.

Krishnaiah, P. R. and Schuurmann, F. J. (1974). On the evaluation of some distributions that arise in simultaneous tests for the equality of the latent roots of the covariance matrix. J. Multivariate Anal., **4**, 265-283.

Krishnaiah, P. R. (1978). Some developments on real multivariate distributions. In: P. R. Krishnaiah, ed., Developments in Statistics, Vol. 1, pp. 135-169. Academic Press, New York.

Krishnaiah, P. R. and Lee, J. C. (1979). On the asymptotic joint distributions of certain functions of the eigenvalues of four random matrices. J. Multivariate Anal., 9, 248-258.

Krishnaiah, P. R. (1980). Computations of some multivariate distributions. In P. R. Krishnaiah, ed., Handbook of Statistics, Vol. 1, pp. 745-971. North-Holland Publishing Company.

Kshirsagar, A. M. (1972). Multivariate Analysis. Dekker, New York.

Lawley, D. N. (1959). Tests of significance in canonical analysis. Biometrika, 46, 59-66.

McKay, R. J. (1976). Simultaneous procedures in discriminant analysis involving two groups. Technometrics, 18, 47-53.

McKay, R. J. (1977). Simultaneous procedures for variable selection in multiple discriminant analysis. Biometrika, 64, 283-290.

Muirhead, R. J. (1978). Latent roots and matrix variates: a review of some asymptotic results. Ann. Statist. 6, 5-33.

Pillai, K. C. S. (1960). Statistical Tables for Tests of Multivariate Hypotheses. Statistical Center, University of Philippines, Manila.

Rao, C. R. (1946). Tests with discriminant functions in multivariate analysis. Sankhya, 7, 407-414.

Rao, C. R. (1966). Covariance adjustment and related problems in multivariate analysis. In P. R. Krishnaiah, ed., Multivariate Analysis, pp. 87-103. Academic Press.

Rao, C. R. (1973). Linear Statistical Inference and Its Applications. John Wiley & Sons, New York.

Schuurmann, F. J., Krishnaiah, P. R. and Chattopadhyay, A. K. (1973). On the distribution of the ratios of the extreme roots to the trace of the Wishart matrix. J. Multivariate Anal., 3, 445-453.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-82-0944	2. GOVT ACCESSION NO. A121 143	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Selection of Variables In Discriminant Analysis		5. TYPE OF REPORT & PERIOD COVERED Technical - June 1982
		6. PERFORMING ORG. REPORT NUMBER 82-15
7. AUTHOR(s) P. R. Krishnaiah		8. CONTRACT OR GRANT NUMBER(s) F49629-82-K-001
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis University of Pittsburgh Pittsburgh, PA 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE 61102F, 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Department of the Air Force Bolling Air Force Base, DC 20332		12. REPORT DATE June 1982
		13. NUMBER OF PAGES 15
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Discriminant analysis, discriminant functions, selection of variables.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper, the author gives a review of some techniques for the selection of variables in discriminant analysis.

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)